Connections between parametric and nonparametric theory

Patrick Breheny

September 4

Score function vs. influence function

- You may have noticed some strong similarities between the influence function and the score function from parametric maximum likelihood estimation
- Letting $U_\theta(x) = \frac{d}{d\theta}\ell(\theta|x)$ denote the score function, in parametric estimation we have

$$\begin{aligned} \mathbb{E}U_\theta(X) &= 0 \\ \mathbb{V}(\hat{\theta}) &\approx \frac{1}{n\mathbb{V}\{U_\theta(X)\}} \\ &= \frac{1}{n\mathbb{E}\{U_\theta^2(X)\}} \end{aligned}$$

- For an observed set of data, $\{U_{\hat{\theta}}(x_i)\}$ are called the *score components*

Score function vs. influence function (cont'd)

- In nonparametric estimation we have

$$\begin{aligned} \mathbb{E}L_F(X) &= 0 \\ \mathbb{V}(\hat{\theta}) &\approx \frac{\mathbb{V}\{L_F(X)\}}{n} \\ &= \frac{\mathbb{E}\{L_F^2(X)\}}{n} \end{aligned}$$

- For an observed set of data, $\{L_{\hat{F}}(x_i)\}$ are called the *influence components*

Influence function of a parametric model

- This connection is not a coincidence; there exists a close relationship between the influence function and the score function of a parametric model
- We can see this relationship directly by deriving the influence function of a parametric model
- *Theorem:* For a parametric model,

$$L_\theta(x) = i(\theta)^{-1} U_\theta(x),$$

  where $i(\theta)$ is the Fisher information
- Thus, the score function and the influence function are scalar multiples of each other, and the multiplication factor is the Fisher information

Parametric estimation of variance

- Note that this reconciles our two definitions:

$$\begin{aligned} \mathbb{V}(\hat{\theta}) &\approx \frac{\mathbb{E}\{L_\theta^2(X)\}}{n} \\ &= \frac{\mathbb{E}\{U_\theta^2(X)\}}{ni(\theta)^2} \\ &= \frac{1}{ni(\theta)} \end{aligned}$$

- Thus, the usual Fisher information method for estimation of variance in a parametric model can be thought of as an influence-function based estimate

Semiparametric estimation of variance

- It is worth noting that $\mathbb{E}\{U_\theta^2(X)\} = i(\theta)$ only if the parametric model is correct
- Maybe it would be a good idea to estimate $\mathbb{V}(\hat{\theta})$ using $n^{-1}\sum_i \hat{L}(x_i)^2$, our variance estimate from the nonparametric delta method
- If we did so, our variance estimate would be

$$\frac{n^{-1}\sum_i L_{\hat{\theta}}(x_i)^2}{n} = \frac{n^{-1}\sum_i U_{\hat{\theta}}(x_i)^2}{ni(\theta)^2},$$

Semiparametric estimation of variance (cont'd)

- In many applications, it turns out that this is indeed a very useful improvement upon the parametric estimation of variance, and provides a consistent estimate for the true variance of $\hat{\theta}$ even if the parametric model is incorrect
- The approach goes by several names:
  - Robust standard errors
  - Semiparametric estimation of variance
  - The "sandwich estimator"
- The last name comes from the vector-based version of the formula:

$$\hat{\mathbb{V}}(\hat{\boldsymbol{\theta}}) = n^{-1} i(\boldsymbol{\theta})^{-1} \left\{ \frac{1}{n} \sum_i U_{\boldsymbol{\theta}}(x_i) U_{\boldsymbol{\theta}}(x_i)^T \right\} i(\boldsymbol{\theta})^{-1}$$